Isolation and structure of repressor-like proteins from the archaeon *Sulfolobus solfataricus*

Co-purification of RNase A with Sso7c[1]

Udo C.T. Oppermann[a,*], Stefan Knapp[b], Valentina Bonetto[a], Rudolf Ladenstein[b], Hans Jörnvall[a]

[a] *Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden*
[b] *Department for Bioscience at Novum, Karolinska Institutet, Center for Structural Biochemistry, S-141 57 Huddinge, Sweden*

Received 13 June 1998; revised version received 26 June 1998

Abstract The thermostable histone-like protein Sso7c (Sso for *Sulfolobus solfataricus*) from the archaeon *Sulfolobus solfataricus* was purified from the supernatant of acid-soluble cell lysates. Reverse phase HPLC of an apparently homogeneous Sso7c protein fraction from Mono S chromatography resulted in resolution of three further peaks. Sequence analysis revealed one of these components to be bovine RNase A, originating from the culture medium and explaining the RNA hydrolyzing activities of Sso7 preparations previously described. Sequence analysis of pure Sso7c showed an ε-Lys methylation pattern identical to that of Sso7d and a single Gln → Glu mutational difference at position 13. The remaining two proteins obtained after HPLC separation were identified as homologues of bacterial repressor-like proteins. Thus, the existence of repressor-like proteins was demonstrated at the protein level in archaea, raising the question of structural and functional consequences of these proteins on the otherwise eukaryotic-like basal transcriptional machinery in archaea.

Key words: Archaeon; Repressor; Transcription

1. Introduction

Living organisms can be divided into three primary domains: Bacteria, Archaea and Eukarya [1]. Archaea, like bacteria, lack a separate nucleus, but in contrast to bacteria, their transcriptional apparatus and DNA organization resemble those of eukaryotes. One striking similarity between archaea and eukaryotes is the presence of true eukaryotic histone homologues in euryarchaea [2,3] proposing a nucleosome-like organization also in archaea. This hypothesis has been supported by electron micrographs of ruptured archaeal nucleoids which reveal the presence of spherical nucleosome-like structures [4,5].

Surprising similarities have been detected also in the basic transcriptional apparatus. RNA polymerases in archaea resemble the eukaryotic enzymes in subunit complexity, sequence and association with initiation factors [6]. In addition, homologues of eukaryotic TATA box binding proteins and the transcription factor TFIIB have been identified in Archaea [7–9]. These archaeal transcription factors can be replaced by human or yeast homologues in in vitro transcription assays indicating similarities also in function [10]. However, whereas a basic understanding of an archaeal transcription mechanism is now emerging, very little is known about the control of this process.

We have chosen the extremely thermophilic crenarchaeon *Sulfolobus solfataricus* for the study of DNA organization and transcription. In *Sulfolobus*, like in other crenarchaea, real eukaryotic histone homologues are lacking. Instead several small and basic DNA binding proteins have been described as histone-like proteins based on their physical properties [11–13]. The most abundant of these proteins is Sso7d which shows no homology to eukaryotic histones and has a β-sheet topology with an SH3 domain folding pattern [14]. Sso7d binds DNA homopolymers with micromolar affinity in a non-cooperative manner and compacts DNA in vitro [15]. In ligase-mediated supercoiling assays binding of Sso7d has been shown to lead to the introduction of negative supercoils [16]. This effect of Sso7d on DNA topology has been recently confirmed by NMR and X-ray studies where significant bending and unwinding of the DNA is observed [17,25]. This report describes the purification and primary structure of repressor-like proteins co-purifying with a homologous Sso7 protein and ribonuclease A.

2. Materials and methods

2.1. Growth and lysis of cells

S. solfataricus, strain DSM 1617 isolated from volcanic hot springs, was obtained from the Deutsche Sammlung für Mikroorganismen (Braunschweig, Germany). Cells were grown aerobically at 75°C in a membrane fermenter (Bioengineering, Switzerland) in a medium [14] containing casamino acids to a maximum density of 1.4 g dry cell weight per l and a constant flow rate of 500 ml medium/h, starting at OD_{556} 1.0. Cells were harvested by centrifugation, resuspended in 10 mM Tris-HCl, pH 8.8, and lysed with a French press.

2.2. Chromatographic procedures

The cell-free lysate from the French press treatment was centrifuged to remove residual debris and the supernatant dialyzed overnight against 100 mM H_2SO_4. After centrifugation and removal of acid precipitated proteins, soluble proteins were dialyzed against 30 mM Tris-HCl, pH 7.4, 100 mM NaCl, and applied to FPLC gel filtration on Superose 6 (90×1 cm), equilibrated in the same buffer, and eluted at a flow rate of 1.0 ml/min. Fractions eluting at 151–158 min were dialyzed against 20 mM potassium phosphate, pH 6.0, applied to FPLC Mono S, and eluted with a stepwise gradient from A (20 mM potassium phosphate, pH 6.0) to B (1 M NaCl in A). Reverse phase HPLC of the fractions obtained was carried out on a C4 stationary phase (Vydac) using 0.1% TFA with a 0–70% acetonitrile gradient in water, after acidification of the samples with formic acid

*Corresponding author. Fax: (46) (8) 33 74 62.
E-mail: udo.oppermann@mbb.ki.se

[1] The 7c2, 7c3 and 7c4 sequences presented in this report are deposited with the EMBL database.

PII: S0014-5793(98)00848-5

to a concentration of 5% (v/v). SDS-PAGE of peptides was carried out using a Tricine based system with 12% polyacrylamide. Staining was achieved with Coomassie brilliant blue.

2.3. Structural analysis

Amino acid compositions were determined with an LKB Alpha Plus analyzer after hydrolysis in evacuated tubes with 6 M HCl, 0.5% phenol. Prior to hydrolysis, peptides were carboxymethylated using iodoacetic acid after reduction with dithiothreitol. Cyanogen bromide cleavage of peptides dissolved in 70% formic acid was performed overnight at room temperature under nitrogen atmosphere. Resulting peptide mixtures were resolved by reverse phase HPLC, analyzed by MALDI-TOF mass spectrometry, or subjected to Edman degradation with an ABI 477A or Milligen Prosequencer 6400/6600. C-terminal sequence analysis was achieved by MALDI-TOF mass spectrometry after ladder generation with carboxypeptidases [18].

3. Results

3.1. Purification of acid-stable, basic proteins from S. solfataricus

Cell lysates of *S. solfataricus*, strain DSM 1617, separated by Mono S (Fig. 1A) revealed a symmetric peak, indicating an apparently homogeneous preparation. However, when separated by C4 reverse phase HPLC with a gradient of 0–70% acetonitrile in 0.1% TFA, four peaks (7c1–7c4) were resolved (Fig. 1B).

3.2. Analysis

Sequence analysis of fraction 7c2 identified it as a homologue (Fig. 2) to the Sso7d protein [14], with one Gln→Glu exchange at position 13, but the same Lys-ε-aminomethylation pattern, confirmed also by mass spectrometry. To exclude an artifactual deamidation by elevated temperature, sequence analysis of recombinant Sso7d treated for 5 days at 85°C was carried out. No deamidation of Gln residues could be observed, indicating that 7c2 is the product of a gene different from that of Sso7d. In contrast, fraction 7c1 revealed identity with positions 27–62 of the bovine RNase A precursor sequence. Sequence analysis of fractions 7c3 and 7c4 (Fig. 2) revealed a close relationship between these two peptides. 7c3 was analyzed for 52 cycles by Edman degradation, for a C-terminal Pro residue by carboxypeptidase digestion to yield a theoretical mass of 5722.6 Da, in good agreement with the mass determined by MALDI-TOF mass spectometry (5725.1 Da). Edman degradation of 7c4 performed for residues 1–48, C-terminal ladder sequence analysis, a total mass of 6156.7 Da determined by MALDI TOF mass spectrometry, showing residues at positions 49–50 to be compatible with DS as in 7c3, resulted in the 7c4 structure as given in Fig. 2.

3.3. Database search and sequence alignment of 7c3 and 7c4 with other DNA binding proteins

Using the sequences obtained, database searches were performed, giving an alignment with similar sequences (Fig. 3). Most of the sequences obtained are open reading frames of bacterial DNA proteins or peptides, constituting repressor proteins, DNA and RNA polymerases or initiation factors. Interestingly, 7c3 and 7c4 align well with two *Escherichia coli* repressor proteins [19], indicating that transcriptional control of gene expression in *S. solfataricus* is performed by 7c3 and 7c4. Included in the alignment are parts of the transcriptional repressor ACCR from *Agrobacterium tumefaciens* [20] and the putative translational initiation factor aIF-1A from *Methanococcus jannaschii* [21].

4. Discussion

An interesting feature of DNA organization in Archaea is the presence of histones which compact DNA into nucleosome-like structures [22]. However, these archaeal histones have so far only been detected in euryarchaeota. Instead, in crenarchaeota a number of 'histone-like proteins' have been described, which share no sequence or structural similarities

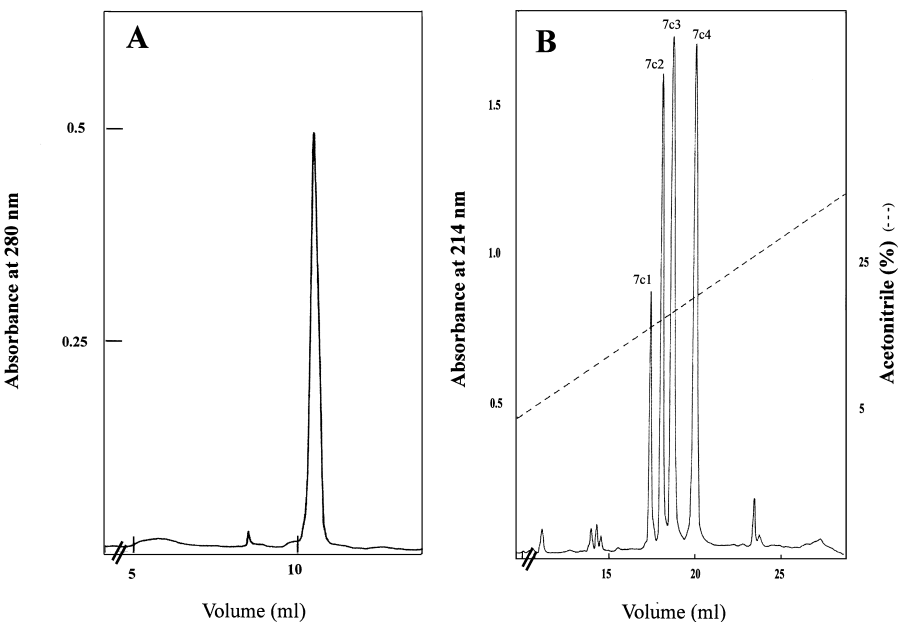

Fig. 1. Mono S cation exchange chromatography (A) of Sso7c from Superose 6 permeation chromatography. The single symmetrical peak in A and the SDS-PAGE pattern indicated homogeneity of the material in A, but the peak pattern in B shows its multicomponent nature. Reverse phase HPLC (B) of the Mono S fraction from A.

Fig. 2. Primary structures of the 7c2, 7c3 and 7c4 proteins from *S. solfataricus*. Arrows pointing right indicate N-terminal sequencer degradation, arrows pointing left C-terminal sequence analysis using carboxypeptidase combined with MALDI-TOF mass spectrometry [18]. ε-Aminomethylation of PTH-lysine residues was detected at the positions asterisked. 7c2 was fragmented by CNBr treatment and resulting peptides were subjected to analysis. Residues 49 and 50 of 7c4 were not recovered in an overlapping fragment. However, based on a mass of 202.8 Da (experimental total mass of 6156.7 Da by MALDI-TOF mass spectrometry versus the mass of sequenced regions), results of the surrounding sequence degradations, and homology with 7c3, the link is concluded to be DS, as shown.

with histones, but show similar general properties of abundance, basicity, acid solubility and small size [23].

To address the questions which proteins participate in the organization of genomic DNA in the thermophilic crenarchaeon *S. solfataricus*, we purified and analyzed the small basic proteins expressed by this organism. As shown earlier, 'histone-like proteins' in *Sulfolobus* can be grouped into different molecular weight classes of 7, 8 and 10 kDa [13], where the 7 kDa family is the most abundant. The most prominent member of this family in *S. solfataricus* is Sso7d, a very basic protein of 63 amino acid residues which consists almost entirely of a β-sheet secondary structure [14]. The biophysical and DNA binding properties of Sso7d have been studied [14,16,24] and its three-dimensional structure in complex with DNA has been elucidated [17]. The crystal structure of the Sso7d homologue in *S. acidocaldarius* (Sac7d) has been determined in complex with DNA showing helix unwinding and a strong bending of DNA, in support of a histone-like function [25]. It seems as if there are two DNA compaction mechanisms in Archaea which differ fundamentally, one like in euryarchaeota involving true histone homologues, and a second, Sso-like mechanism in crenarchaeota.

Apart from Sso7d, we were able to purify a strictly homologous protein, which differs from Sso7d at only one position, 13, with Glu instead of Gln. Since no deamidation of this Gln occurred in vitro, it is reasonable to assume that Sso7c is not the result of deamidation in vivo. In agreement with this conclusion two strictly homologous genes have been identified in the related organism *S. acidocaldarius* [26]. According to the original nomenclature of Sso molecules we propose the name Sso7c for this protein since it is less basic than Sso7d. In the NMR model of a Sso7d-DNA complex, residue Gln-13 does not participate in DNA binding. Thus, an exchange at this position is not likely to interfere with DNA binding.

The Sso7d protein is heterogeneously methylated on five of the 14 ε-aminolysine residues [12]. None of the methylated lysine residues participates in DNA binding [27]. In contrast to the Sso homologues Sac7d and Sac7e (Sac for *S. acidocaldarius*), which differ in methylation [26], Sso7d and Sso7c are methylated at identical positions.

An RNase activity has been assigned to Sso7c [28]. However, as now shown, Sso7c, considered to be pure, can be separated into four major components by C4 reverse phase HPLC, 7c1, 7c2, 7c3 and 7c4 (in order of elution) and sequence analysis of 7c1 revealed 100% identity to bovine RNase A. We therefore conclude that RNase A is a contam-

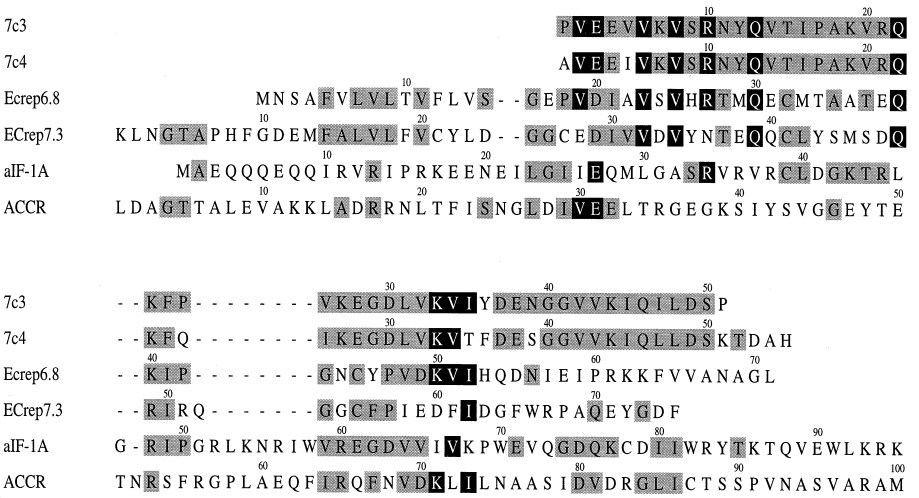

Fig. 3. Alignment of 7c3 and 7c4 with bacterial repressor proteins from *E. coli* (Ecrep6.8 and Ecrep7.3; SwissProt accession numbers P21419 and P29010 [19]), translation initiation factor aIF-1A from *M. jannaschii* (SwissProt accession number MJ0445) and the repressor protein ACCR from *A. tumefaciens* (SwissProt accession number P32104 [20]). Regions with more than 33% identity are marked in gray, with more than 66% identity in black.

ination from the casamino acid medium. Thus our results provide a reasonable explanation for the intriguing observation that Sso7 should function as an RNase [28].

Apart from Sso7c, two other peptides were isolated (7c3 and 7c4, Figs. 1 and 2). They exhibit similarities with two repressor proteins (Ecrep 6.8 and Ecrep 7.3, Fig. 3) of *E. coli* [19], the repressor protein ACCR in *A. tumefaciens* [20] and the initiation factor aIF-1A from the archaeon *M. jannaschii* [21]. Because of the high similarities to these transcriptional repressors, we assume a regulatory role in gene transcription for the 7c3 and 7c4 proteins in Archaea. Very little is known of this process in Archaea and the two proteins now isolated appear to be the first repressor-like proteins described in this kingdom. However, repressor-like transcriptional regulation has been suggested based on palindromic sequences upstream of TATA boxes in phage ΦH lysogens of *Halobacterium halobium* [29] as well as on a repressor binding site in *Methanococcus maripaludis* [30]. Putative repressors have also been suggested from genome structures of *M. jannaschii* [21], but the proteins have not been isolated.

Acknowledgements: This study was supported by EC biotechnology grants (BIO4-CT97-2123, and Biotechnology of Extremophiles), the Swedish Medical Research Council (13X-3532), the Swedish Cancer Society (project 1806) and the Deutsche Forschungsgemeinschaft. Expert assistance by Ella Cederlund and Irene Byman is gratefully acknowledged as is help with figures by Carina Palmberg.

References

[1] Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Proc. Natl. Acad. Sci. USA 87, 4576–4579.
[2] Sandman, K., Krzychi, J.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1990) Proc. Natl. Acad. Sci. USA 87, 5788–5791.
[3] Starich, M.R., Sandman, K., Reeve, J.N. and Summers, M.F. (1996) J. Mol. Biol. 255, 187–203.
[4] Bohrmann, B., Kellenberger, E., Arnold-Schulz-Gahmen, B., Sreenivas, K., Suryanarayana, T., Stroup, D. and Reeve, J.N. (1994) J. Struct. Biol. 112, 70–78.
[5] Shioda, M.B., Sugimori, K., Shiroya, T. and Takayanagi, S. (1989) J. Bacteriol. 171, 4514–4517.
[6] Baumann, P., Qureshi, S.A. and Jackson, S.P. (1995) Trends Genet. 11, 279–283.
[7] Qureshi, S.A., Baumann, P., Rowlands, T., Khoo, B. and Jackson, S.P. (1995) Nucleic Acids Res. 23, 1775–1781.
[8] Ouzounis, C. and Sander, C. (1992) Cell 71, 189–190.
[9] Creti, R., Londei, R. and Cammarano, P. (1993) Nucleic Acids Res. 21, 2942.
[10] Wettach, J., Gohl, H.P., Tschochner, H. and Thomm, M. (1995) Proc. Natl. Acad. Sci. USA 92, 472–476.
[11] Choli, T., Wittmann-Liebold, B. and Reinhard, R. (1988) J. Biol. Chem. 263, 7087–7093.
[12] Choli, T., Henning, P., Wittmann-Liebold, B. and Reinhard, R. (1988) Biochim. Biophys. Acta 950, 193–203.
[13] Grote, M., Dijk, J. and Reinhard, R. (1986) Biochim. Biophys. Acta 873, 405–413.
[14] Baumann, H., Knapp, S., Lundbäck, T., Ladenstein, R. and Härd, T. (1995) Nature Struct. Biol. 1, 808–819.
[15] Lurz, R., Grote, M., Dijk, J., Reinhardt, R. and Dobrinski, B. (1986) EMBO J. 5, 3715–3721.
[16] Lopez-Gracia, P., Knapp, S., Forterre, P. and Ladenstein, R. (1998) Nucleic Acids Res. (in press).
[17] Agback, P., Baumann, H., Knapp, S., Ladenstein, R. and Härd, T. (1998) J. Biol. Chem. (submitted).
[18] Bonetto, V., Bergman, A.C., Jörnvall, H. and Sillard, R. (1997) J. Protein Chem. 16, 371–374.
[19] Cam, K., Bejar, S.M., Gil, D. and Bouche, J.P. (1988) Nucleic Acids Res. 16, 6327–6338.
[20] Beck von Bodman, S., Hayman, T.G. and Farrand, S.K. (1992) Proc. Natl. Acad. Sci. USA 89, 643–647.
[21] Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M. and Venter, J.C. (1996) Science 273, 1058–1073.
[22] Reeve, J.N., Sandman, K. and Daniels, C.J. (1997) Cell 89, 999–1002.
[23] Schmid, M.B. (1990) Cell 63, 451–453.
[24] Knapp, S., Karshikoff, A., Berndt, K.D., Christova, P., Atanasov, B. and Ladenstein, R. (1996) J. Mol. Biol. 264, 1132–1144.
[25] Robinson, H., Gao, Y.-G., McCrary, B.S., Edmondson, S.P., Shriver, J.W. and Wang, A.H.-J. (1998) Nature 392, 202–205.
[26] McAfee, J.G., Edmondson, S.P., Datta, P.K., Shriver, J.W. and Gupta, R. (1995) Biochemistry 34, 10063–10077.
[27] Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R. and Härd, T. (1995) J. Mol. Biol. 247, 840–846.
[28] Fusi, P., Tedeschi, G., Aliverti, G., Ronchi, S., Tortora, P. and Guerritore, A. (1993) Eur. J. Biochem. 211, 305–310.
[29] Ken, R. and Hackett, N.R. (1991) J. Bacteriol. 173, 955–960.
[30] Cohen-Kupiec, R., Blank, C. and Leigh, J.A. (1997) Proc. Natl. Acad. Sci. USA 94, 1316–1320.